Effective DGA Family Classification using a Hybrid Shallow and Deep Packet Inspection Technique on P4 Programmable Switches

Ali AlSabeh\textsuperscript{1}, Kurt Friday\textsuperscript{2}, Jorge Crichigno (Presenter)\textsuperscript{1}, Elias Bou-Harb\textsuperscript{2}

\textsuperscript{1}University of South Carolina, SC
\textsuperscript{2}The University of Texas at San Antonio, TX

IEEE International Conference on Communications (IEEE ICC)
May 30, 2023 - Rome, Italy
Introduction

• Attackers often use a Command and Control (C2) server to establish communication between infected host/s and bot master

• Domain Generation Algorithms (DGAs) are the de facto dynamic C2 communication method used by malware, including botnets, ransomware, and many others\(^1\)

---

DGA Attacks

• DGAs evade firewall controls by frequently changing the domain name selected from a large pool of candidates
• The malware makes DNS queries to resolve the IP addresses of these generated domains
• Only a few of these queries will be successful; most of them will result in Non-Existent Domain (NXD) responses

(1) DNS queries. (2) (NXD) replies. (3) Eventually, a query for the actual domain is sent and malware-C2 communication starts.
DGA Attacks

- DGAs evade firewall controls by frequently changing the domain name selected from a large pool of candidates.
- The malware makes DNS queries to resolve the IP addresses of these generated domains.
- Only a few of these queries will be successful; most of them will result in Non-Existent Domain (NXD) responses.
Existing Mitigation Strategies

• Approaches rely on contextual network traffic analysis (context-aware) or domain name analysis, without considering network traffic (context-less)

• Most research efforts focus on DGA detection, i.e., they perform binary classification in order to segregate DGAs from benign traffic

• In addition to DGA detection, it is helpful to classify DGA malware based on the family (Trojan, Backdoor, etc.)
Motivation

• Context-aware approaches analyze the network traffic behavior to fingerprint DGAs
  ➢ Slow since they typically analyze batches of traffic offline

• Domain-name (context-less) approaches obtain high accuracy with ML models
  ➢ The use of a general-purpose CPU/GPU may create a bottleneck due to high traffic volume

• There is a need for a system that uses both context-aware and context-less features to classify DGAs
Contribution

• Proposing a novel P4 scheme that uses a hybrid context-aware and context-less feature extraction technique entirely in the data plane

• Implementing Deep Packet Inspection (DPI) on Intel’s Tofino ASIC that extracts and analyzes domain names within 3 microseconds

• Evaluating the proposed approach on 50 DGA families collected by crawling GBs of malware samples

• Highlighting the effectiveness of the proposed work in terms of accuracy, performance
Overview P4 Switches

• P4 switches permit the programmer to program the data plane
  ➢ Customized packet processing
  ➢ High granularity in measurements
  ➢ Per-packet traffic analysis and inspection
  ➢ Stateful memory processing

```p4
136 /******************************************** PARSER ********************************************/
137
def state parse_ethernet {
138     packet.extract(hdr.ethernet);
139     transition select(hdr.ethernet.etherType) {
140         Type_IPV4: parse_ipv4;
141         default: accept;
142     }
143 }
144
def state parse_ipv4 {
145     packet.extract(hdr.ipv4);
146     verify(hdr.ipv4.ihl <= 5, error.IPHdrTooShort);
147     transition select(hdr.ipv4.ihl) {
148         5 : accept;
149         default : parse_ipv4_option;
150     }
151 }
152 }
```
Overview P4 Switches

- P4 switches permit the programmer to program the data plane
  - Customized packet processing
  - High granularity in measurements
  - Per-packet traffic analysis and inspection
  - Stateful memory processing
- If the P4 program compiles, it runs on the chip at line rate

Reproduced from N. McKeown. Creating an End-to-End Programming Model for Packet Forwarding. Available: [https://www.youtube.com/watch?v=fiBuao6YZI0&t=4216s](https://www.youtube.com/watch?v=fiBuao6YZI0&t=4216s)
Proposed System

• The P4 PDP switch collects and stores the context-aware features of the hosts

• When an NXD response is received, the switch performs DPI on the domain name to extract domain features

• The switch sends the collected features to the control plane

• The control plane runs the intelligence to classify the DGA family and initiate the appropriate incidence response
Proposed System

• Context-aware features
  ➢ For each host in the network, the following features are stored in the data plane:
    ▪ Number of IP addresses contacted
    ▪ Inter-arrival Time (IAT) between such IP packets
    ▪ Number of DNS requests made
    ▪ Time it takes for the first NXD response to arrive
    ▪ IAT between subsequent NXD responses
  ➢ Collected in the data plane
Proposed System

• Context-less features
  - It computes the bigram of the domain name; a bigram model may suffice to predict whether a domain name is a legitimate human readable domain
    \[
    \text{score}(d) = \sum_{\forall \text{ subdomain } s \in d} \left( \sum_{\forall \text{ bigram } b \in s} f_s^b \right)
    \]
    Where \( f_s^b \) is the frequency of the bigram \( b \) in the subdomain \( s \)
  - The frequency value of a bigram \( b \) is pre-computed and stored in a Match-Action Table (MAT)
  - The lower the score, the more random the domain name
  - Example: the bigrams of “google” are: “$g”, “go”, “oo”, “og”, “gl”, “le”, “e$”
Evaluation

• Dataset
  ➢ Hundreds of GB of malware samples; 1,311 samples containing 50 DGA families
  ➢ To collect DGA-based malware, only samples that receive NXD responses containing domain names generated by DGAs (based on DGArchive\(^1\)) are considered

• Experimental setup
  ➢ The collected dataset was used to train ML models offline on a general-purpose CPU
  ➢ 80% of data was used for training and 20% for testing

Evaluation

- Accuracy (Acc), F1 score, and Precision (Prec) of different ML classifiers during the first 8 NXD responses received were reported.
- The Random Forest (RF) model performed best.
  - The Accuracy (Acc) starts at 92% from the first NXD response received and reaches 98% by the 8th NXD response.

<table>
<thead>
<tr>
<th>NXD count</th>
<th>RF Acc</th>
<th>RF F1</th>
<th>RF Prec</th>
<th>SVM Acc</th>
<th>SVM F1</th>
<th>SVM Prec</th>
<th>MLP Acc</th>
<th>MLP F1</th>
<th>MLP Prec</th>
<th>LR Acc</th>
<th>LR F1</th>
<th>LR Prec</th>
<th>GNB Acc</th>
<th>GNB F1</th>
<th>GNB Prec</th>
</tr>
</thead>
<tbody>
<tr>
<td>NXD 1</td>
<td>0.923</td>
<td>0.907</td>
<td>0.902</td>
<td>0.872</td>
<td>0.856</td>
<td>0.847</td>
<td>0.87</td>
<td>0.843</td>
<td>0.829</td>
<td>0.716</td>
<td>0.679</td>
<td>0.667</td>
<td>0.726</td>
<td>0.688</td>
<td>0.688</td>
</tr>
<tr>
<td>NXD 2</td>
<td>0.951</td>
<td>0.943</td>
<td>0.943</td>
<td>0.899</td>
<td>0.893</td>
<td>0.893</td>
<td>0.904</td>
<td>0.897</td>
<td>0.9</td>
<td>0.76</td>
<td>0.741</td>
<td>0.747</td>
<td>0.727</td>
<td>0.701</td>
<td>0.707</td>
</tr>
<tr>
<td>NXD 3</td>
<td>0.964</td>
<td>0.958</td>
<td>0.964</td>
<td>0.918</td>
<td>0.913</td>
<td>0.914</td>
<td>0.924</td>
<td>0.914</td>
<td>0.912</td>
<td>0.767</td>
<td>0.74</td>
<td>0.743</td>
<td>0.649</td>
<td>0.668</td>
<td>0.732</td>
</tr>
<tr>
<td>NXD 4</td>
<td>0.966</td>
<td>0.961</td>
<td>0.963</td>
<td>0.906</td>
<td>0.905</td>
<td>0.912</td>
<td>0.916</td>
<td>0.909</td>
<td>0.915</td>
<td>0.79</td>
<td>0.765</td>
<td>0.758</td>
<td>0.633</td>
<td>0.635</td>
<td>0.692</td>
</tr>
<tr>
<td>NXD 5</td>
<td>0.97</td>
<td>0.966</td>
<td>0.967</td>
<td>0.915</td>
<td>0.91</td>
<td>0.911</td>
<td>0.919</td>
<td>0.91</td>
<td>0.907</td>
<td>0.77</td>
<td>0.735</td>
<td>0.746</td>
<td>0.604</td>
<td>0.615</td>
<td>0.689</td>
</tr>
<tr>
<td>NXD 6</td>
<td>0.975</td>
<td>0.972</td>
<td>0.973</td>
<td>0.914</td>
<td>0.911</td>
<td>0.912</td>
<td>0.922</td>
<td>0.915</td>
<td>0.918</td>
<td>0.794</td>
<td>0.767</td>
<td>0.783</td>
<td>0.617</td>
<td>0.627</td>
<td>0.716</td>
</tr>
<tr>
<td>NXD 7</td>
<td>0.977</td>
<td>0.976</td>
<td>0.979</td>
<td>0.92</td>
<td>0.915</td>
<td>0.915</td>
<td>0.929</td>
<td>0.924</td>
<td>0.93</td>
<td>0.799</td>
<td>0.771</td>
<td>0.78</td>
<td>0.61</td>
<td>0.613</td>
<td>0.714</td>
</tr>
<tr>
<td>NXD 8</td>
<td>0.98</td>
<td>0.979</td>
<td>0.981</td>
<td>0.917</td>
<td>0.912</td>
<td>0.914</td>
<td>0.93</td>
<td>0.923</td>
<td>0.921</td>
<td>0.764</td>
<td>0.73</td>
<td>0.735</td>
<td>0.631</td>
<td>0.618</td>
<td>0.65</td>
</tr>
</tbody>
</table>

RF: Random Forest; SVM: Support Vector Machine; MLP: Multilayer perceptron; LR: Logistic Regression; GNB: Gaussian Naive Bayes
Evaluation

- Feature extraction time of the proposed approach and EXPLAIN
- EXPLAIN’s available source code was tested on a general-purposed CPU with 64 GB RAM, 2.9 GHz processor with 8 cores
Conclusion and Discussion

• In this work, we propose a hybrid feature extraction technique relying on context-aware and context-less features to classify DGA families.

• Context-aware features characterize the network traffic behavior of the DGAs and require shallow packet inspection (no degradation to the throughput).

• Context-less features study the statistical and structural characteristics of the domain names relating to NXDs using DPI.

• With 50 DGA families analyzed, the proposed approach achieves 92% accuracy with RF classifier from the first NXD response and reaches up to 98% by the 8th NXD response.

• We plan to explore other techniques that are robust against encrypted DNS traffic, in addition to collecting more DGA families.
Acknowledgement

• Thanks to the National Science Foundation (NSF)

• Activities in the CI Lab at the UofSC are supported by NSF, Office of Advanced Cyberinfrastructure (OAC), awards 2118311 and 2104273
References


Thank You

• Contact info for further questions
  ➢ aalsabeh@email.sc.edu
  ➢ jcrichigno@cec.sc.edu

• CyberInfrastructure Lab (CI Lab) website
  ➢ http://ce.sc.edu/cyberinfra/